CONVERGENCE OR DIVERGENCE OF EXPERT MENTAL MODELS: THE UTILITY OF KNOWLEDGE STRUCTURE ASSESSMENT IN TRAINING RESEARCH

Stephen M. Fiore
Jennifer Fowlkes
Laura Martin-Milham
University of Central Florida

Randall L. Oser
Naval Air Warfare Center
Training Systems Division

In this paper we discuss the impact of differing knowledge structure measurement techniques on assessing instructor mental models for behaviors associated with Situation Awareness. Our goals were, first, to investigate the degree to which an expert model for such behaviors actually exists, and second, to determine the degree to which experts, varying along a number of dimensions, assess these behaviors using differing knowledge structure measurement techniques. The results show substantial agreement in concept relatedness across differing measures, but less agreement across differing expert groups. Our discussion focuses on the differing measures and their ability to assess the knowledge structures associated with experts differing in their training roles and we review the implications of these findings for training researchers.

Acquiring the knowledge structures necessary for task expertise is at the root of all training programs, yet a clear understanding of the nature of expertise remains elusive. Part of the problem is that expertise can manifest itself in any number of ways (e.g., Ericsson & Charness, 1994; Patel, Drury, & Shalin, 1998), and we explore differences in the way experts may represent task-relevant knowledge. In particular, we explore whether an expert model actually exists within a given domain (e.g., Britton & Tidwell, 1995; Rowe, Cooke, & Rivera, 1998; Shanteau, 1998).

We were interested in a complex cognitive task (i.e., low-level navigation); specifically, expert assessment of the relation between generic Situation Awareness behaviors associated with this task. This data was collected in the context of a larger investigation involving event-based situation awareness assessment using the SALIANT (Situation Awareness Linked Indicators Adapted to Novel Tasks) behavioral indicators (Muniz, Stout, Bowers, & Salas, 1998). In addition to exploring the representational nature of the Situation Awareness behaviors, we sought to determine whether these behaviors would be differentially assessed dependent upon either the experience of the instructor (i.e., pilot or navigator), or the knowledge structure measurement technique (i.e., Card Sort or Pathfinder analyses). Thus, if one's perspective differently influences how one perceives these behaviors, given that the instructors under study vary in their training experience, differences in knowledge representation may be exhibited. As such, the purpose of this research was not to evaluate the efficacy of the knowledge structure measures as an indicator of performance. Rather, the knowledge structure measures were used as an index with which to gauge and compare expert understanding of the Situation Awareness behaviors.

Current Study

Over the past decade, there has been substantial methodological progress in assessing knowledge structures (e.g., Glaser, 1989; Nichols, Chipman, & Brennan, 1995; Schvaneveldt, 1990). Nonetheless, much debate exists about the nature of mental models and their relation to knowledge structure assessment techniques (e.g., Jonassen, Beissner, & Yacci, 1993; Nichols, et al. 1995). Most conceptualize knowledge structures in terms of a particular pattern or relationship among a given set of information (e.g., facts, procedures, diagrams) and use a given assessment method to ascertain a subject's mastery of a given domain, that is, their mental model (see, for example, Hoffman, 1992).

The purpose of the present study was to compare two differing knowledge structure assessment techniques among experienced aviation instructors. Specifically, our first goal was to determine whether we could identify an organizing framework with which to group the SALIANT behaviors. Such a grouping may be used to guide training programs. In particular, such a framework may have wide-spread implications for training system design and performance measurement and feedback. Thus, if instructors exhibit marked agreement in the nature of their groupings, this would suggest possible methods for organizing the behaviors which may in turn facilitate their acquisition and retention.

Our second goal was to determine how the knowledge structure assessment techniques may interact with instructor experience and may influence the resultant representations. We hypothesized that, first, the differing techniques (Card Sorts vs. Pathfinder similarity ratings) may differently measure aspects of the expert knowledge (cf. Dorsey, Campbell, Foster, & Miles, 1999). In particular, we hypothesized that the knowledge structures generated from similarity ratings may result in differing patterns among the experts than those generated from Card Sorts. Second, we hypothesized that, although the participants were drawn from the same population (i.e., aviation instructors), given their differing training, experience, and roles (navigator versus pilot), their expert models may diverge with respect to these Situation Awareness behaviors. Thus, despite the fact that the instructors may share relatively the same level of experience within a domain (i.e., number of flight hours), their unique backgrounds may impact their expert model.

Methods

Participants

Participants were 13 military aviators from a Naval aviation training squadron. Eight of the participants were T-34 instructor pilots and five were T-39 instructor pilots. The primary role of the T-34 group was that of pilot and the

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2000	**N/A**	**-**

4. TITLE AND SUBTITLE	
Convergence Or Divergence Of Expert Mental Models: The Utility Of Knowledge Structure Assessment In Training Research	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Central Florida**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**4**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

primary role of the T-39 group was that of navigator. The T-34 group had an average of 1,880 flight hours and the T-39 group had an average of 1,717 flight hours. This difference in number of flight hours was not significant ($F < 1$).

Stimuli

Participants were presented with 16 concepts associated with Situation Awareness behaviors. These concepts were derived from the SALIANT behavioral inventory (Muniz, Stout, Bowers, & Salas, 1998).

Procedure

For the Card Sort task, participants were presented with 16 index cards on which the concepts were typed. They were instructed to group these behaviors into as many or as few categories as they desired. For the similarity rating task, Pathfinder software was used to elicit similarity ratings among the set of concepts (see Schvaneveldt, 1990, for a discussion of the Pathfinder algorithms).

Results

In order to assess the relation among concept pairs, for the Card Sort task, each possible concept pair (N = 120) was coded with a 0 if the participant did not group them in the same category, or a 1 if they were grouped in the same category. An analogous coding can be derived from the Pathfinder similarity ratings output. Thus, we were able to compare these two measures in order to determine whether the instructors make similar concept pairings when presented with two somewhat distinct tasks.

To facilitate discussion of our findings, the results section is divided into two parts. In order to assess the validity of the SALIANT concepts used in the training evaluation, in the first section we discuss the data from the Card Sort and Pathfinder tasks in relation to the SALIANT behaviors overall. For the purposes of investigating the degree to which experts varying in background converge or diverge in agreement, in the second section we discuss how differences across the expert groups are manifested depending upon the measure being used.

SALIANT Behaviors

As mentioned, our purpose here was to determine whether we could identify an organizing framework with which to group the SALIANT behaviors. For this analysis, each concept pair received a mean rating computed across subjects, but separately for the Card Sort and the Pathfinder tasks. Thus, concept pairs could have a score ranging from 0 (if no participant ever grouped that pair), to 1 (if all participants grouped that pair). Our first hypothesis was that the differing knowledge structure assessment techniques would result in different representations of the expert knowledge. Thus, we were interested in the degree to which the differing measures correlated in their assessment of the concept pairs. Over all participants, a significant correlation ($r = .64$, $p < .001$, $df = 118$) was found between the Pathfinder and Card Sort ratings. This failure to support the hypothesis suggests that the measures were producing relatively equal ratings for the concept pairs.

In order to assess the nature of the groupings that the instructors were making, a qualitative analysis of the mean concept pair ratings was conducted. We determined which concept pairs were consistently being grouped together in both the Card Sort and the Pathfinder data. For this analysis we considered only those pairs that had a mean rating in the 75th percentile (translating to approximately over 50% of the participants considering these pairs related). Based upon this analysis, notably similar groupings were found across the Card Sort and Pathfinder data. Table 1 lists the categories that were gleaned from this analysis.

Comparisons Across Expert Groups

Rather than only looking for similarities in assessment, we additionally analyzed the degree to which experts may differ in their representation of these behaviors. We hypothesized that, although the participants are drawn from the same population, given their differing operational roles (navigator versus pilot), their expert models may diverge with respect to these Situation Awareness behaviors. Using the concept pair coding scheme described above we computed correlations with all possible participant pairings. Additionally, based upon comparisons across the Pathfinder nets, a similarity index was computed for each participant pairing. This similarity index is derived from the number of common links in pairs of pathfinder nets and it is based upon the proportion of the links in either network that are in both networks.

Table 1. *SALIANT Behavioral Categories based upon Card Sort and Pathfinder Similarity Ratings.*

SA Category	SALIANT Behavior
Spatial Orientation	Spatial awareness
	Uses available information
	Cross checks information
	Scans VFR/IFR
Cue Sharing	Provides and requests backup
	Reports problems
	Informs others of action taken
Problem Solving	Locates potential sources of problems
	Resolves discrepancies
	Anticipates consequences
Information Management	Provides info in advance
	Uses standard communication format
	Briefs status
Task Management	Takes action at appropriate time
	Knowledge of task
	Skilled time sharing among tasks

Overall, both knowledge structure measurement techniques yielded substantial agreement among the group of experts. Specifically, for each participant pairing, the mean Pathfinder net similarity index was .22 (a relatively high rating, see, for example, Schvaneveldt, 1990). Additionally, the Pathfinder inter-subject correlations (derived from the actual concept pair similarity ratings) yielded a significant mean correlation of .28. With the Card Sort concept pairing correlations, the experts were in somewhat less agreement on the Situation Awareness behaviors, with the overall correlation of .15 being only marginally significant ($p < .06$, 1-

tailed). Nonetheless, the focus of this aspect of the investigation was on the degree to which expert agreement differed depending upon varying basis for their expertise. Thus, in the following sections we compare the differing levels of agreement based upon two criteria: primary instructor role (T-34 pilot vs. T-39 navigator); and number of flight hours (more experience vs. less experience).

Instructor Role as Comparison Basis

For the purposes of comparing across the differing instructors, three comparison levels were derived. Specifically, the first comparison level consisted of the data from each T-34 instructor being compared with each other (yielding 28 possible comparisons from the 8 T-34 participants), the second consisted of each T-39 instructor being compared with each other (yielding 10 possible comparisons from the 5 T-39 participants), and the third consisted of the T-34 instructors being compared with the T-39 instructors (yielding 40 possible comparisons from the combination of participants). Thus, for each participant pair type we created the following variables for analysis: a mean similarity index from the Pathfinder nets, a mean inter-subject correlation based upon Pathfinder similarity ratings, and a mean inter-subject correlation based upon the Card Sort groupings.

For the similarity index, there was a main effect of Pilot Pairing $F(2, 77) = 4.99$, $p < .01$. Post Hoc tests showed that the mean similarity index for within T-34 pairings (M = .24), was significantly higher than the within T-39 pairings (M = .18), and the T-34/T-39 pairings (M = .21). In order to compare the Pathfinder and Card Sort inter-subject correlations, the data was subjected to a 2x3 mixed-model ANOVA with Correlation Type (Pathfinder vs. Card Sort) as the within participant variable and Plane Pairing Type (T34/T34, T39/T39, and T34/T39), as the between groups variable. This analysis yielded a significant effect for Correlation Type, $F(1, 75) = 18.88$, $p < .001$, with the mean Pathfinder correlations (M = .28) being significantly greater than the mean Card Sort correlations (M = .15). Additionally, there was a significant interaction between Correlation Type and Plane Pairing Type, $F(2, 75) = 6.82$, $p < .01$ (refer to Figure 1). Post Hoc tests showed that each of the mean inter-subject correlations with the Pathfinder data were significantly different. There were no significant differences for the Card Sort inter-subject correlations.

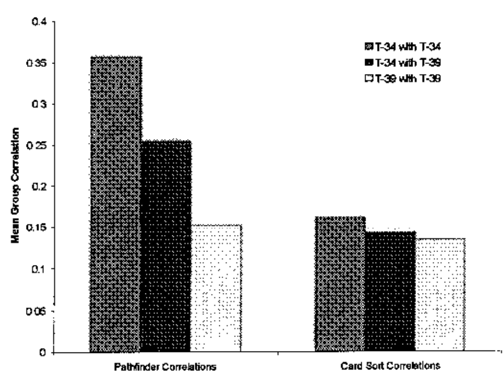

Figure 1. Mean inter-subject correlation for the differing role comparisons.

Number of Flight Hours as Comparison Basis

We next compared across the differing instructors, but with a more commonly used metric (number of flight hours). Three additional comparison levels were derived and were based upon contrasting more experienced pilots (M = 2,142 hours) to less experienced pilots (M = 1,540 hours). Thus, even though all participants were experienced enough to be instructors, a substantial range in flight hours existed. The first comparison level consisted of each more experienced instructor being compared with each other (yielding 15 possible comparisons from these 6 participants), the second consisted of each less experienced instructor being compared with each other (yielding 21 possible comparisons from these 7 participants), and the third consisted of the more and less experienced instructors being compared with each other (yielding 42 possible comparisons from the combination of participants). For each of these participant pair types we created the variables identical to those used in the previous analysis.

No effect was found for the similarity index ($F < 1$). The mean similarity rating for the more experienced pairings (M = .21) was no different than the less experienced pairings (M = .22) and the more/less pairings (M = .21). In order to compare the Pathfinder and Card Sort inter-subject correlations, the data was subjected to a 2x3 mixed-model ANOVA with Correlation Type (Pathfinder vs. Card Sort) as the within participant variable and Participant Pairing Type (more with more, less with less, more with less), as the between groups variable. This analysis yielded a marginally significant interaction between Correlation Type and Experience Level Type, $F(2, 75) = 2.79$, $p < .07$ (refer to Figure 2). Post Hoc tests showed that only the Card Sort mean inter-subject correlations for the more/more group were significantly different from the more/less and less/less groups. There were no significant differences for the Pathfinder inter-subject correlations.

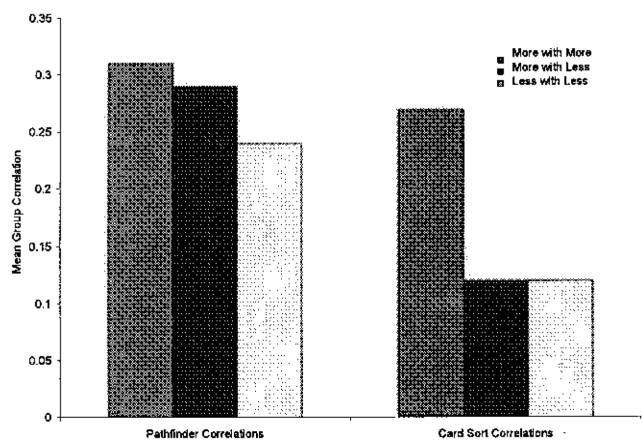

Figure 2. Mean inter-subject correlation for the differing flight hour comparisons.

Discussion

This investigation found that the general Situation Awareness behaviors used in this training evaluation show a high level of agreement across the sample of instructors.

Additionally, differing knowledge structure assessment techniques converged to suggest that these SALIANT behaviors can be grouped according to concrete dimensions. For example, concepts associated with spatial orientation consistently were grouped together in differing knowledge structure measures. Thus, using two differing techniques of knowledge elicitation, a potential organizing framework with which to group the SALIANT behaviors was identified. This framework, once validated with a larger sample, may be applied in training system design and assist in performance measurement and feedback.

We also found that aviators with different experience and roles may view these behaviors somewhat differently. In particular, the T-34 community was in significantly greater agreement than the T-39 community when the Pathfinder similarity rating correlations and similarity index were used. No differences were found when the Card Sort correlation was used. Conversely, when number of flight hours was the criteria, the Card Sort correlations were able to reliably distinguish among the differing groups, while the Pathfinder data found no differences. These findings suggest the following: 1) samples from differing communities of experts may be required when the task-relevant knowledge is general in nature (e.g., SA behaviors); and, 2) that caution is warranted when only a single knowledge structure assessment technique is used to assess mental models. Specifically, multiple methods may be necessary to converge on a clearer understanding of the manner in which the sample population represents the information in question.

Last, given the substantial difference in the ease with which these differing knowledge structure measures are administered, the marked similarity in the resulting patterns of conceptual groupings is noteworthy. In particular, when assessing the data overall, because the Card Sort method yielded a virtually identical pattern as the Pathfinder method, this suggests that Card Sorts may be a more efficient manner of eliciting certain forms of expert knowledge. Although we acknowledge that Pathfinder analyses result in a greater depth of data, researchers should a priori determine whether the relative benefit of the additional data outweighs the cost (e.g., increases in time to administer associated with increases in number of concepts).

Despite these findings, several caveats are warranted. First, although the number of experts used was relatively adequate, a larger sample size may more accurately determine the degree to which these techniques and populations lead to convergence or divergence. Second, even though the two measures showed substantial agreement for the concept pairings, the number of concepts used was relatively low. Therefore, when the number of concepts in question is greater, the Card Sort methodology and Pathfinder ratings may actually diverge. Clearly, additional research is warranted given the ubiquity of these techniques in applied and basic investigations of mental models.

References

Britton, B. K. & Tidwell, P. (1995). Cognitive structure testing: A computer system for diagnosis of expert-novice differences. In P. D. Nichols, S. F. Chipman, & R. L. Brennan. *Cognitively diagnostic assessment* (pp. 251-278). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorsey, D. W., Campbell, G. E., Foster, L. L., & Miles, D. E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance, 12,* 31-57.

Ericsson, K. A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725-747.

Glaser, R. (1989). Expertise in learning: How do we think about instructional processes now that we have discovered knowledge structure? In D. Klahr & D. Kotosfky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 269-282). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jonassen, D. H., Beissner, K. & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Muniz, E., Stout, R., Bowers, C., & Salas, E. (1998). A methodology for measuring team situational awareness: Situational linked indicators adapted to novel tasks (SALIANT). The First Annual Symposium/Business Meeting of the Human Factors & Medicine on Collaborative Crew Performance in Complex Systems. Edinburg, United Kingdom.

Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). *Cognitively diagnostic assessment.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Patel, S. C., Drury, C. G., & Shalin, V. L. (1998). Effectiveness of expert semantic knowledge as a navigational aid within hypertext. *Behaviour & Information Technology, 17,* 313-324.

Rowe, A. L., Cooke, N. J., & Rivera, K. (1998). Assessing knowledge structures: An analysis of knowledge standards. Poster presented at the 42nd Annual Meeting of the Human Factors and Ergonomics Society, Chicago, IL.

Shanteau, J. (1998). Why do experts disagree with each other? Paper presented at the 4th Conference on Naturalistic Decision Making, Airlie, VA.

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization.* Norwood, NJ: Ablex.